Gestion des bases de données

DU Chef de Projet en Recherche Clinique
23/11/2012

Fabrice GOURMELON
URC/CIC Necker - Cochin

Introduction

A. Qu’est-ce qu’une donnée ?
B. Qu’est-ce qu’une base de données ?
C. Définition de base de données
D. Accès à une base de données
E. Système de Gestion des Bases de Données (SGBD)
F. Outils utilisés en Recherche clinique
G. Data Management

A. Qu’est-ce qu’une donnée ?

- C’est une information quelconque
  - Je m’appelle Fabrice
- Relation entre plusieurs informations
  - J’enseigne les bases de données au DU Chef de Projet à Paris

Collection ordonnée d’information pour un même individu

B. Qu’est-ce qu’une base de données

- Ensemble volumineux d’informations
- Caractéristique essentielle : mémoriser les informations de manière permanente

C. Définition

- Une Base de données est un gros ensemble d’informations structurées mémorisées sur un support permanent.

D. Accès à une base de données

- Données réparties en tables (entités)
  - Une table est composée de lignes
  - Une ligne est un ensemble fixe de champs (attributs)
  - Un champ à un nom et un type

- Chaque table est un fichier stocké sur mémoire secondaire

D. Accès à une base de données (2)

- Recours à un outil chargé de :
  - Gérer les fichiers constituant une base de données
  - Prendre en charge les fonctionnalités de protection et de sécurité
  - Fournir les différents types d'interfaces nécessaires à l'accès aux données

E. Système de Gestion des Bases de Données (SGBD)

- Définition
  - Un Système de Gestion de Bases de Données (SGBD) est un logiciel qui permet de manipuler les informations stockées dans une base de données tout en cachant la complexité des opérations

E. Système de Gestion des Bases de Données (SGBD) (2)

- Un ensemble de services pour :
  - Contrôle simple de l'accès aux données
  - Accès aux informations à de multiples utilisateurs
  - Manipuler les données présentes dans la base (insertion, suppression, modification)
- Se décompose en trois sous-systèmes:
  - Gestion de fichiers pour le stockage des informations sur le support physique
  - Gestion de l'ordonnancement des informations
  - Interface avec l'utilisateur

E. Système de Gestion des Bases de Données (SGBD) (3)

- Exemples de SGBD
  - Microsoft Access
  - Microsoft SQL Server
  - MySQL
  - OpenOffice.org Base
  - Oracle Database
  - PostgreSQL
  - Sybase

F. Outils utilisés en Recherche Clinique

- Microsoft Access
- CleanWEB (MySQL)
- Clinsight (Oracle Database)
- Oracle Clinical (Oracle Database)
- Clintrial (Oracle Database)
- OpenClinica (PostgreSQL)

Utilisation :

CRF papier + Saisie dans une base de données

CRF électronique

G. Data Management

- Définition
  - Le Data Management est l'activité de gestion des données de la recherche clinique permettant d'obtenir une base de données cohérente et exploitable par les statisticiens afin de répondre aux objectifs de l'étude.

I. Préparation de la base de données

A. Annotation du CRF
B. Dictionnaire des données
C. Création des tables
D. Schéma relationnel
E. Création du masque de saisie
F. Définition des contrôles de cohérence
G. Convention de saisie / Mode d'emploi
H. Test de la base

Collaboration de plusieurs intervenants du Projet (Investigateurs, Chef de Projets, ARC, Data Manager, Biostatisticien)

A. Annotation du CRF

- Première étape pour structurer l'information du questionnaire dans une base de données
  - Traduire « informatiquement » le CRF
- Fractionner le CRF en groupes de données
  - Groupe de données => Table
- Annoter un CRF vierge en indiquant un nom de variable pour chaque items à saisir
  - Nom et type des variables, Liste des codes

I. Préparation de la base de données

Fabrice Gourmelon - URC/CIC Necker - Cochin

A. Annotation du CRF (2)

Patient : |__|__|-|__|__|__|-|__|__|-|__| Centre : |__|__|

VISITE

isite 2

DTE_VISITE
Date de la visite |__|__|/|__|__|/|__|__|__|__|

INTERROGATOIRE

MODIF_TTT_ET
Y-a-t-il eu une modification dans la prise du traitement à l'étude : O Oui (1) O Non (0)

APP_EI
Y-a-t-il eu apparition d'événements indésirables depuis la dernière visite O Oui (1) O Non (0)

MODIF_TTT_CONC
Y-a-t-il eu une modification dans la prise de traitement concomitants : O Oui (1) O Non (0)

EXAMCLIN

EXAMEN CLINIQUE

ETAT_GEN
Etat général
O Excellent (1)
O Bon (2)
O Moyen (3)
O Mauvais (4)

POIDS
Poids (kg) |__|__|__|, |__|

TEMP
Température (°C) |__|__|, |__|

PAS
PAS (mmHg) |__|__|__|


Table


Variable


Codes

B. Dictionnaire des données

- Description plus complète de la variable
  - Table où elle se situe
  - Nom de la variable
  - Type de la variable (texte, date, entier, réel…)
  - Format de la variable (taille, décimale…)
  - Unité
  - Libellé
  - Liste de codes (pour les QCM)
  - Clé (primaire, étrangère)

Doit être revu par le Data Manager et le Statisticien

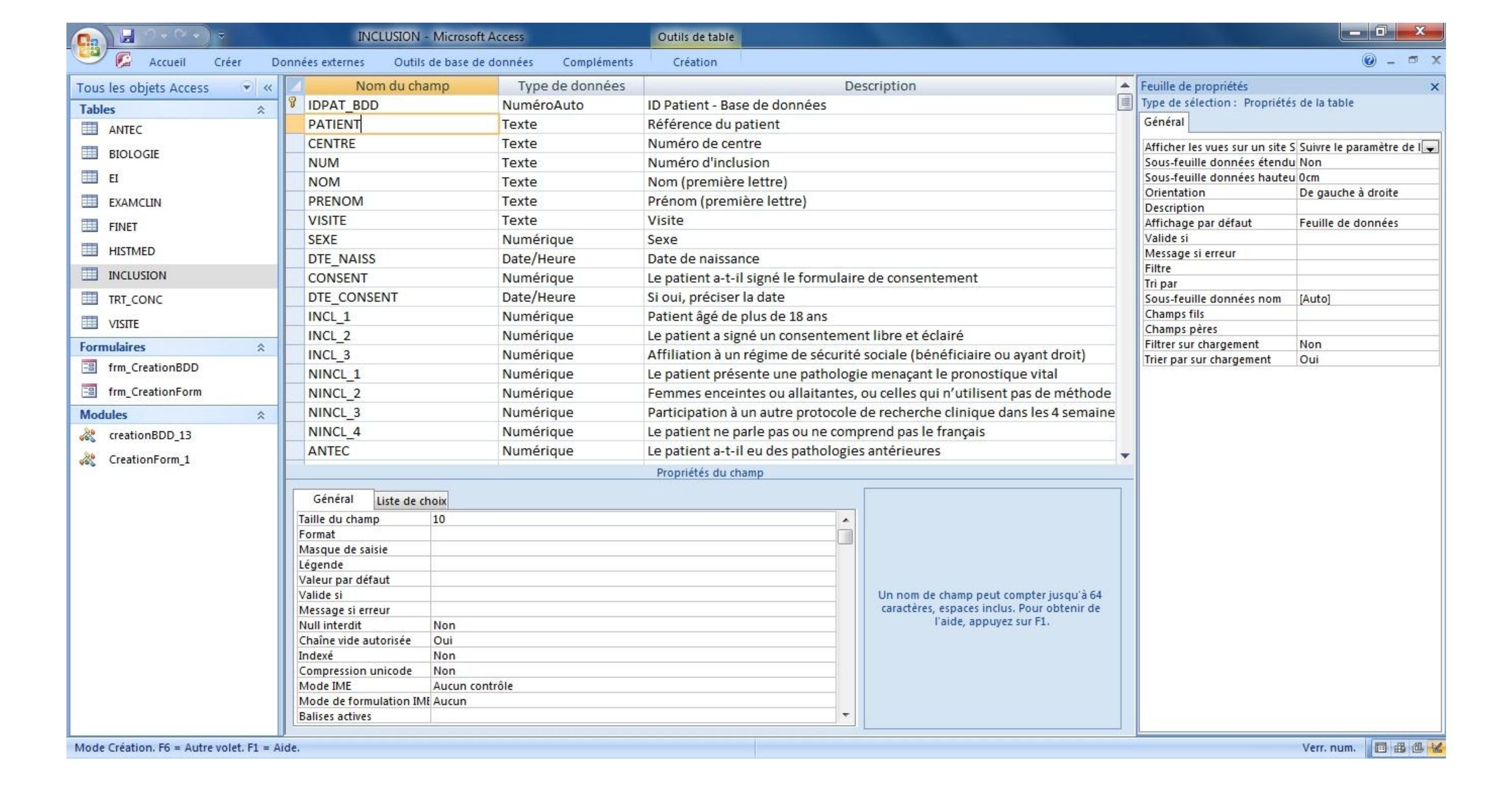
B. Dictionnaire des données (2)

	A	B	C	D	E	F	G	H	I
1	TABLE	VARIABLE	TYPE	FORMAT_SAS	UNITE	LABEL	CODES	CLE	NUM
33	HISTMED	DESC_PREM_SYMP	texte	$150.		*Description des premiers symptômes*		0	32
34	HISTMED	EVOL_MALADIE	texte	$150.		*Evolution de la maladie*		0	33
35	VISITE	IDVISITE_BDD	entier_auto	3.		*ID VISITE - Base de données*		1	34
36	VISITE	IDPAT_BDD	entier	3.		*ID Patient - Base de données*		3	35
37	VISITE	VISITE	texte	$50.		*Visite*	"VISITE1";"VISITE2"	0	36
38	VISITE	DTE_VISITE	date	DDMMYY10.		*Date de la visite*		0	37
39	VISITE	MODIF_TTT_ET	qcm_1	oui_non.		*Y-a-t-il eu une modification dans la prise du traitement à l'étude*	1;"Oui";0;"Non"	0	38
40	VISITE	APP_EI	qcm_1	oui_non.		*Y-a-t-il eu apparition d'événements indésirables depuis la dernière visite*	1;"Oui";0;"Non"	0	39
41	VISITE	MODIF_TTT_CONC	qcm_1	oui_non.		*Y-a-t-il eu une modification dans la prise de traitement concomitants*	1;"Oui";0;"Non"	0	40
42	VISITE	BIOLOGIE	qcm_1	oui_non.		*Le bilan biologique a-t-il été réalisé*	1;"Oui";0;"Non"	0	41
43	VISITE	BIOLOGIE_NON	texte	$150.		*Si non, préciser la raison*		0	42
44	EXAMCLIN	IDEXAMCLIN_BDD	entier_auto	3.		*ID EXAMCLIN - Base de données*		1	43
45	EXAMCLIN	IDVISITE_BDD	entier	3.		*ID VISITE - Base de données*		2	44
46	EXAMCLIN	ETAT_GEN	qcm_1	etat_gen.		*Etat général*	1;"Excellent";2;"Bon";3;"Moyen";4;"Mauvais"	0	45
47	EXAMCLIN	POIDS	reel	3.1	Kg	*Poids*		0	46
48	EXAMCLIN	TEMP	reel	2.1	°C	*Température*		0	47
49	EXAMCLIN	PAS	entier	3.	mmHg	*PAS*		0	48
50	EXAMCLIN	PAD	entier	3.	mmHg	*PAD*		0	49
51	EXAMCLIN	FC	entier	3.	/min	*Fréquence cardiaque*		0	50
52	BIOLOGIE	IDBIOLOGIE_BDD	entier_auto	3.		*ID BIOLOGIE - Base de données*		1	51
53	BIOLOGIE	IDVISITE_BDD	entier	3.		*ID VISITE - Base de données*		2	52
54	BIOLOGIE	DTE_PRELEV_BIO	date	DDMMYY10.		*Date du prélèvement*		0	53
55	BIOLOGIE	HEMO	reel	2.1	g/dL	*Hémoglobine*		0	54
56	BIOLOGIE	HT	reel	3.1	%	*Hématocrite*		0	55
57	BIOLOGIE	VGM	reel	3.1	μm^3	*VGM*		0	56
58	BIOLOGIE	CCMH	reel	2.1	g/dL	*CCMH*		0	57
59	BIOLOGIE	TCMH	reel	2.1	pg	*TCMH*		0	58
60	BIOLOGIE	RETICU	entier	6.	/mm^3	*Réticulocytes*		0	59

C. Création des tables

- Début de la construction de la base de données
- Création des Tables dans le SGBD à partir du dictionnaire de données
  - Création manuelle dans le SGBD
  - Utilisation du langage SQL
  - Création automatique à partir du dictionnaire de données

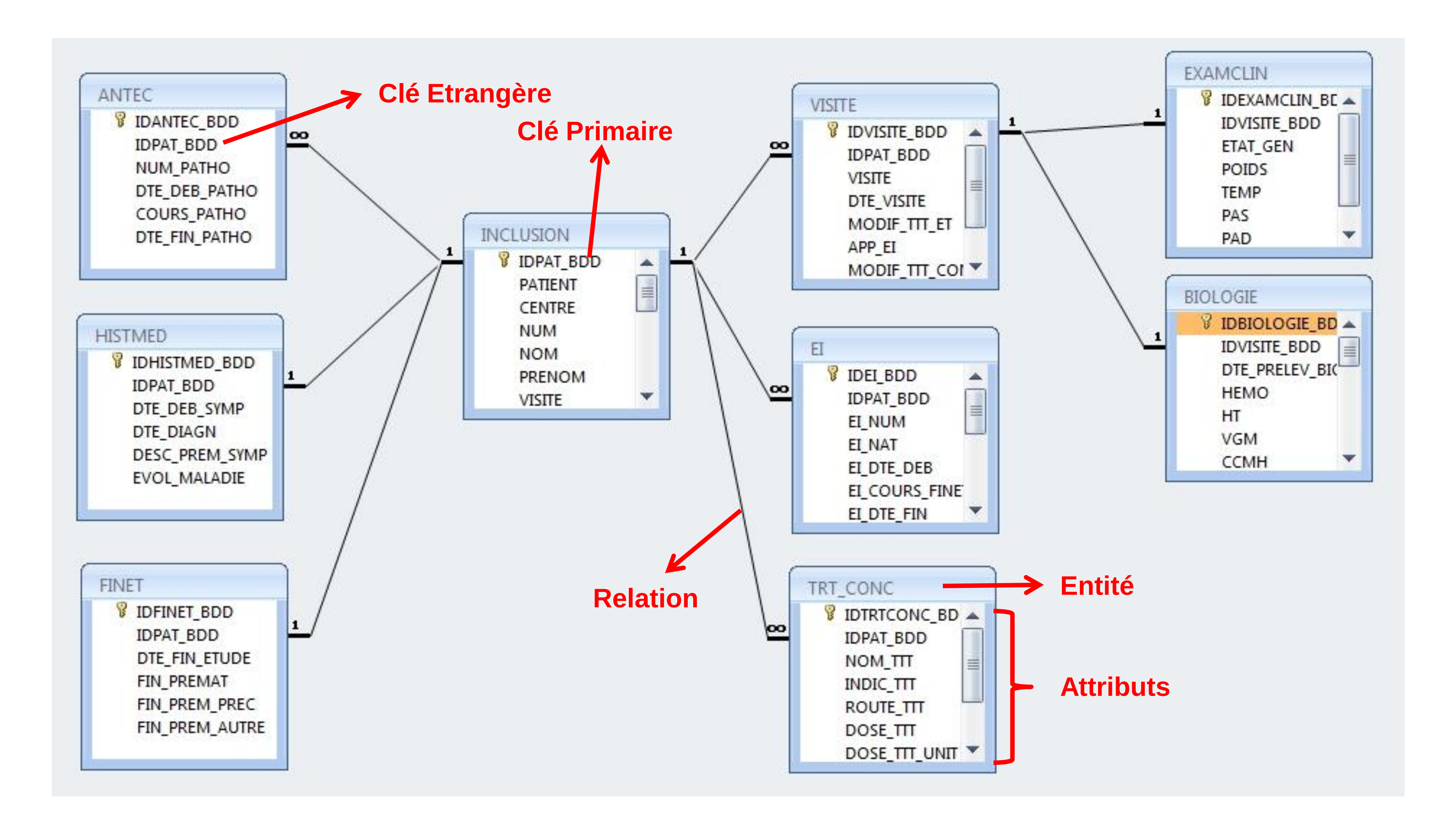
C. Création des tables (2)



D. Schéma relationnel

- Ensemble des schémas des tables définies dans cette base de données
  - **Clés primaires** (Identifiant unique d'une ligne de la table)
  - **Clés étrangères**
    - Création de liaisons fixes entre les tables
- Une clé étrangère dans une table est une clé primaire dans une autre table
  - Relation : Un-à-Un, Un-à-Plusieurs
  - Intégrité référentielle

D. Schéma relationnel (2)

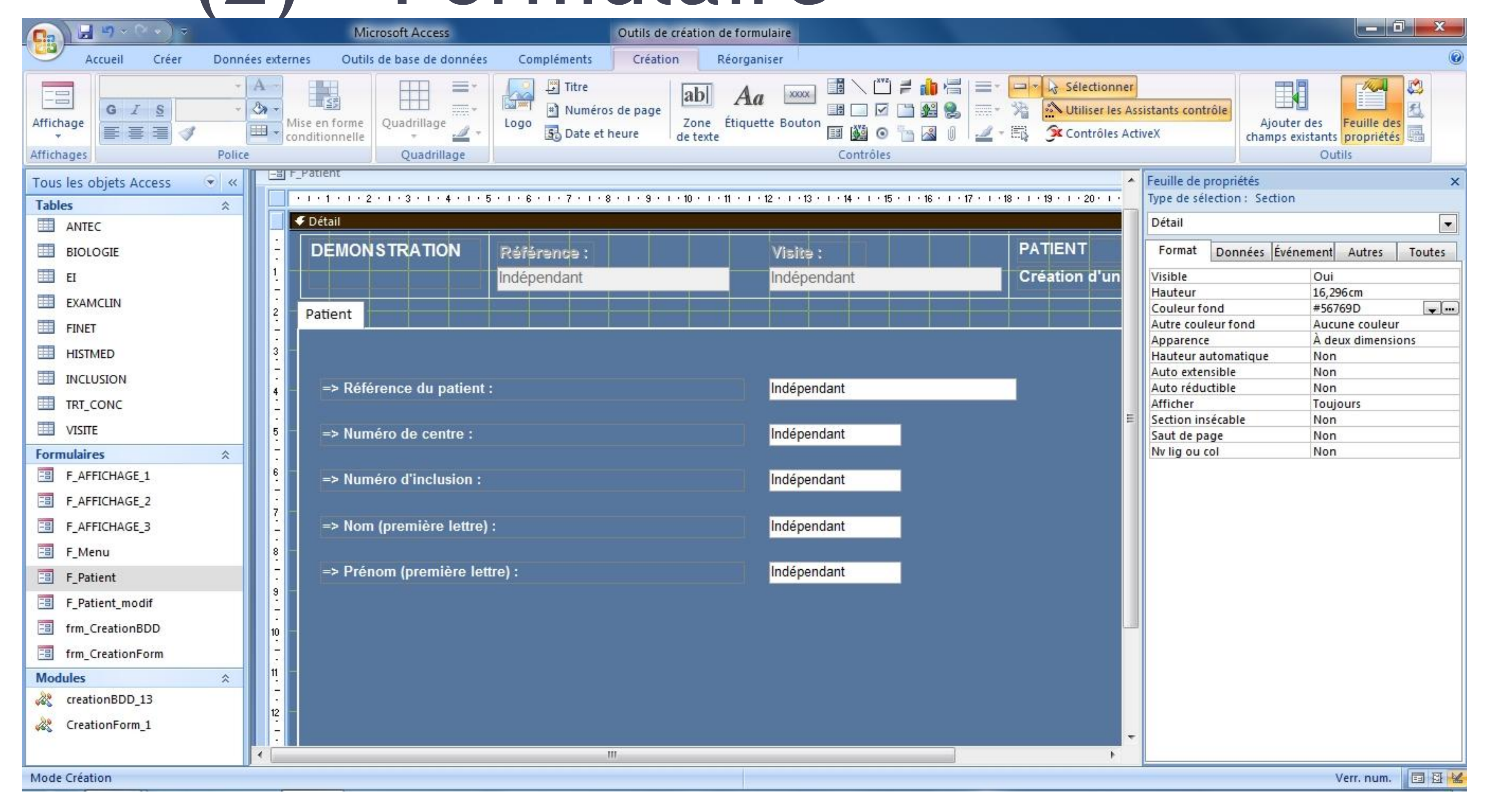


E. Création du masque de saisie

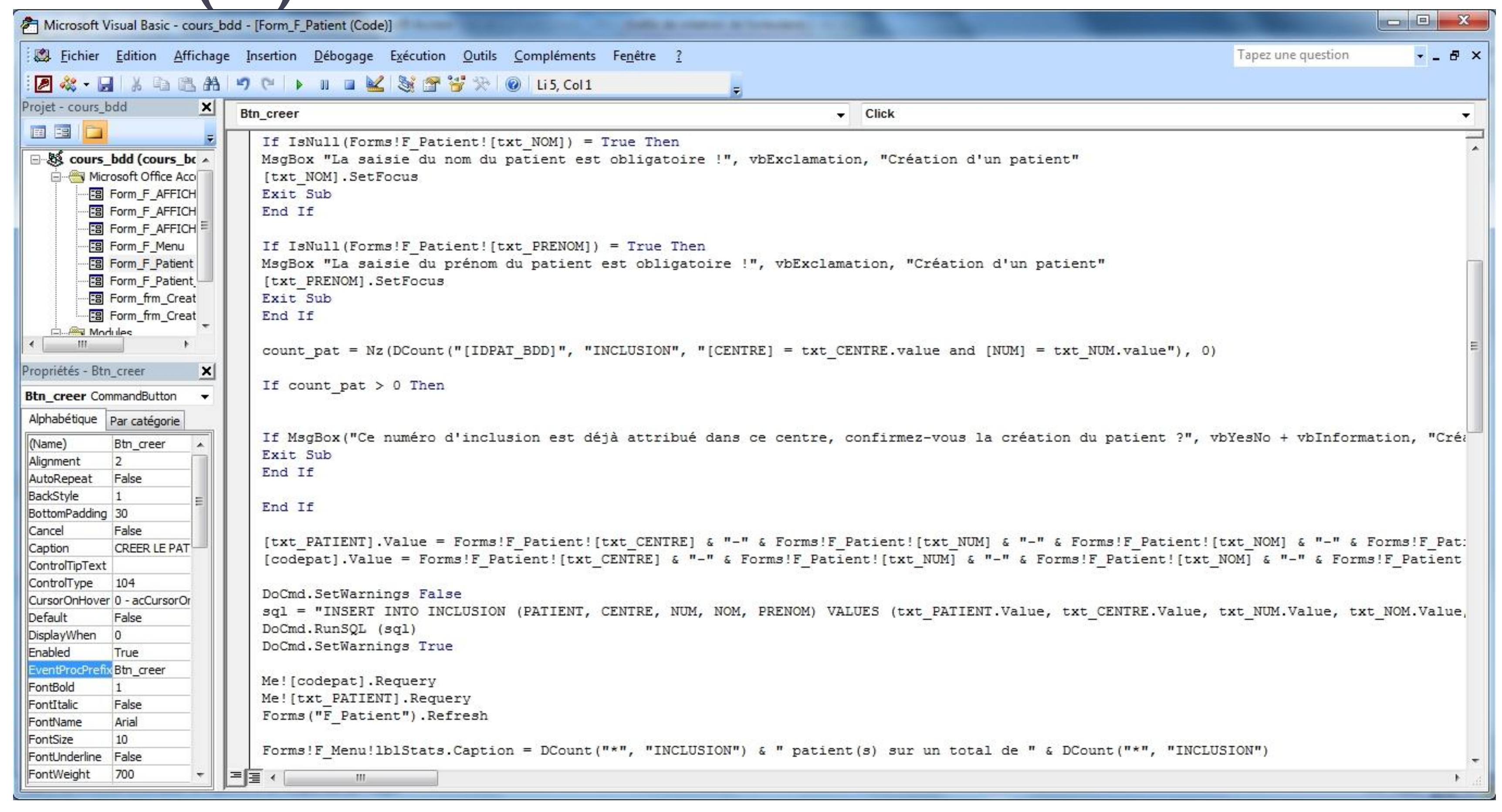
- Espaces de saisie dans l'interface utilisateur comportant plusieurs zones
  - Texte, Radio-button, Combo-box, Cases à cocher, Listes, Boutons
- Interactivité entre l’utilisateur et la base de données
- Homogénéisation du format des entrées
- Contrôle des entrées

Fidèle au CRF Papier, Conviviale

E. Création du masque de saisie (2) - Formulaire



E. Création du masque de saisie (3) – Code VBA




```
If IsNull(Forms!F_Patient![txt_NOM]) = True Then
MsgBox "La saisie du nom du patient est obligatoire !", vbExclamation, "Création d'un patient"
[txt_NOM].SetFocus
Exit Sub
End If

If IsNull(Forms!F_Patient![txt_PRENOM]) = True Then
MsgBox "La saisie du prénom du patient est obligatoire !", vbExclamation, "Création d'un patient"
[txt_PRENOM].SetFocus
Exit Sub
End If

count_pat = Nz(DCount("[IDPAT_BDD]", "INCLUSION", "[CENTRE] = txt_CENTRE.value and [NUM] = txt_NUM.value"), 0)

If count_pat > 0 Then


If MsgBox("Ce numéro d'inclusion est déjà attribué dans ce centre, confirmez-vous la création du patient ?", vbYesNo + vbInformation, "Créa
Exit Sub
End If

End If

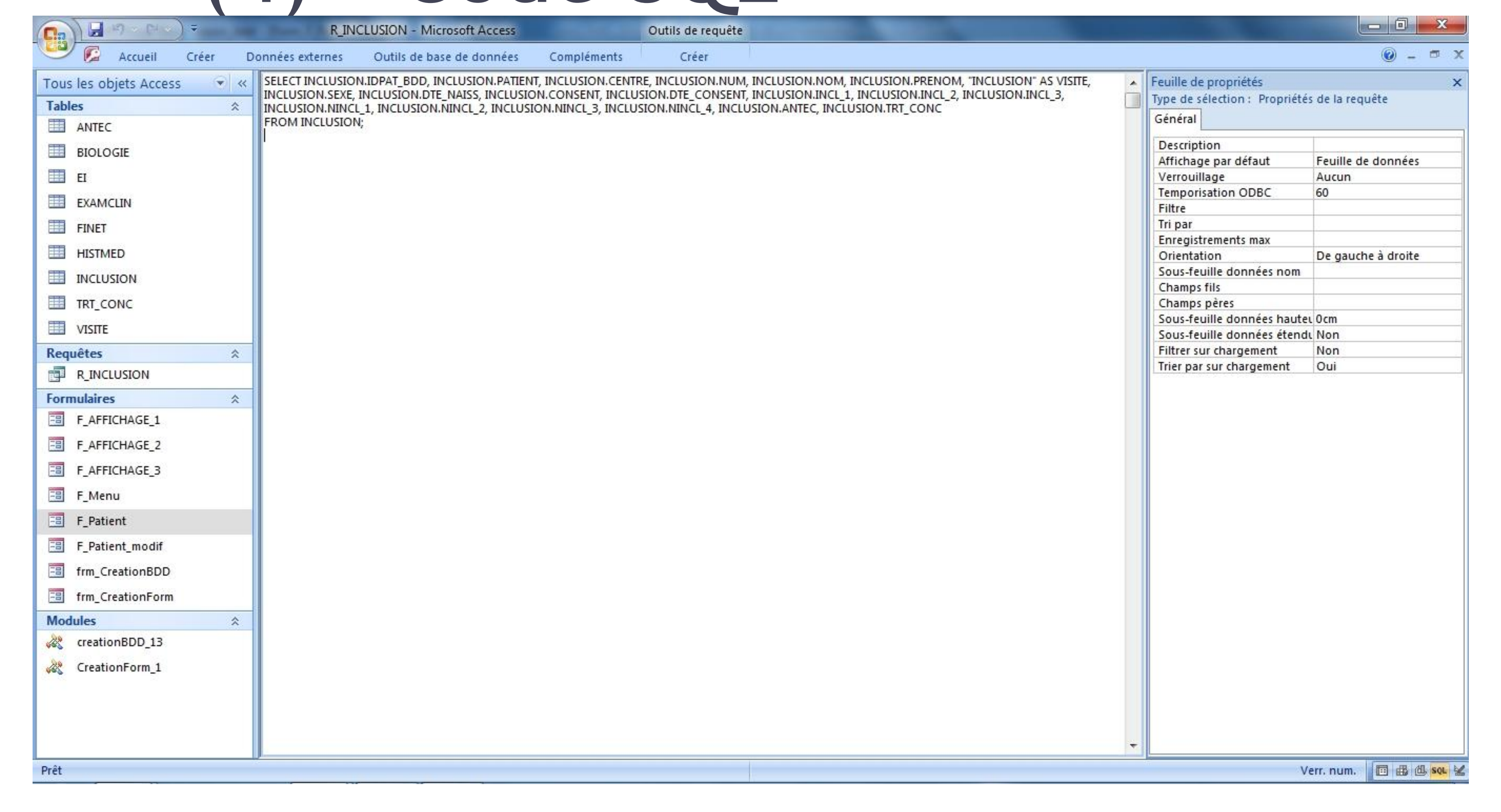
[txt_PATIENT].Value = Forms!F_Patient![txt_CENTRE] & "-" & Forms!F_Patient![txt_NUM] & "-" & Forms!F_Patient![txt_NOM] & "-" & Forms!F_Pat:
[codepat].Value = Forms!F_Patient![txt_CENTRE] & "-" & Forms!F_Patient![txt_NUM] & "-" & Forms!F_Patient![txt_NOM] & "-" & Forms!F_Patient

DoCmd.SetWarnings False
sql = "INSERT INTO INCLUSION (PATIENT, CENTRE, NUM, NOM, PRENOM) VALUES (txt_PATIENT.Value, txt_CENTRE.Value, txt_NUM.Value, txt_NOM.Value,
DoCmd.RunSQL (sql)
DoCmd.SetWarnings True

Me![codepat].Requery
Me![txt_PATIENT].Requery
Forms("F_Patient").Refresh

Forms!F_Menu!lblStats.Caption = DCount("*", "INCLUSION") & " patient(s) sur un total de " & DCount("*", "INCLUSION")
```

E. Création du masque de saisie (4) – Code SQL



E. Création du masque de saisie (5) - Résultat

DEMONSTRATION | Référence : 01-001-A-B | Visite : INCLUSION | PATIENT Création d'un patient

Patient | Données démographiques | Critères d'inclusion | Critères de non inclusion | Pathologie / Traitement

=> Référence du patient : 01-001-A-B

=> Numéro de centre : 01

=> Investigateur recruteur : Dr TESTEUR

=> Numéro d'inclusion : 001

=> Nom (première lettre) : A

=> Prénom (première lettre) : B

VALIDER | Annuler

F. Définition des contrôles de cohérences

- Liste des contrôles effectués sur les données
  - Référence du contrôle
  - Table et variable concernée
  - Message pour les investigateurs

- Réalisation à partir des documents suivants
  - Protocole
  - CRF
  - Dictionnaire des données

Fabrice Gourmelon - URC/CIC Necker - Cochin

F. Définition des contrôles de cohérences (2)

1	EC_ID	EC_REF	EC_TABLE	EC_VARIABLE	EC_LABEL	EC_QUERY	EC_TEST
11	10	BO1	BIOLOGIE	LYMPH	Lymphocytes (/mm^3)	Donnée hors bornes [200 - 10000], Confirmez-vous?	BIOLOGIE.LYMPH < 200 OR BIOLOGIE.LYMPH > 10000
12	11	BO1	BIOLOGIE	MONO	Monocytes (/mm^3)	Donnée hors bornes [50 - 3000], Confirmez-vous?	BIOLOGIE.MONO < 50 OR BIOLOGIE.MONO > 3000
13	12	BO1	BIOLOGIE	PLAQ	Plaquettes (/mm^3)	Donnée hors bornes [70000 - 900000], Confirmez-vous?	BIOLOGIE.PLAQ < 70000 OR BIOLOGIE.PLAQ > 900000
14	13	BO1	BIOLOGIE	NA	Sodium (mmol/L)	Donnée hors bornes [60 - 300], Confirmez-vous?	BIOLOGIE.NA < 60 OR BIOLOGIE.NA > 300
15	14	BO1	BIOLOGIE	K	Potassium (mmol/L)	Donnée hors bornes [1 - 10], Confirmez-vous?	BIOLOGIE.K < 1 OR BIOLOGIE.K > 10
16	15	BO1	BIOLOGIE	CL	Chlore (mmol/L)	Donnée hors bornes [40 - 250], Confirmez-vous?	BIOLOGIE.CL < 40 OR BIOLOGIE.CL > 250
17	16	BO1	BIOLOGIE	HCO3	Bicarbonates (mmol/L)	Donnée hors bornes [10 - 80], Confirmez-vous?	BIOLOGIE.HCO3 < 10 OR BIOLOGIE.HCO3 > 80
18	17	BO1	BIOLOGIE	CA	Calcium (mmol/L)	Donnée hors bornes [1 - 10], Confirmez-vous?	BIOLOGIE.CA < 1 OR BIOLOGIE.CA > 10
19	18	BO1	BIOLOGIE	ASAT	ASAT (UI/L)	Donnée hors bornes [5 - 100], Confirmez-vous?	BIOLOGIE.ASAT < 5 OR BIOLOGIE.ASAT > 100
20	19	BO1	BIOLOGIE	ALAT	ALAT (UI/L)	Donnée hors bornes [5 - 100], Confirmez-vous?	BIOLOGIE.ALAT < 5 OR BIOLOGIE.ALAT > 100
21	20	BO1	BIOLOGIE	GGT	GGT (UI/L)	Donnée hors bornes [5 - 120], Confirmez-vous?	BIOLOGIE.GGT < 5 OR BIOLOGIE.GGT > 120
22	21	BO1	BIOLOGIE	CREAT	Créatinine (µmol/L)	Donnée hors bornes [20 - 250], Confirmez-vous?	BIOLOGIE.CREAT < 20 OR BIOLOGIE.CREAT > 250
23	22	BO1	EXAMCLIN	POIDS	Poids (Kg)	Donnée hors bornes [15 - 300], Confirmez-vous?	EXAMCLIN.POIDS < 15 OR EXAMCLIN.POIDS > 300
24	23	BO1	EXAMCLIN	TEMP	Température (°C)	Donnée hors bornes [35 - 45], Confirmez-vous?	EXAMCLIN.TEMP < 35 OR EXAMCLIN.TEMP > 45
25	24	BO1	EXAMCLIN	PAS	PAS (mmHg)	Donnée hors bornes [90 - 250], Confirmez-vous?	EXAMCLIN.PAS < 90 OR EXAMCLIN.PAS > 250
26	25	BO1	EXAMCLIN	PAD	PAD (mmHg)	Donnée hors bornes [30 - 130], Confirmez-vous?	EXAMCLIN.PAD < 30 OR EXAMCLIN.PAD > 130
27	26	BO1	EXAMCLIN	FC	Fréquence cardiaque (/min)	Donnée hors bornes [30 - 200], Confirmez-vous?	EXAMCLIN.FC < 30 OR EXAMCLIN.FC > 200
28	27	BO1	TRT_CONC	DOSE_TTT	Dose totale journalière	Donnée hors bornes [0 - 9999,9], Confirmez-vous?	TRT_CONC.DOSE_TTT < 0 OR TRT_CONC.DOSE_TTT > 9999,9
29	28	DA1	ANTEC	DTE_DEB_PATHO	Date de début	Date inférieure à : INCLUSION.DTE_NAISS, Confirmez-vous?	ANTEC.DTE_DEB_PATHO < INCLUSION.DTE_NAISS
30	29	DA1	ANTEC	DTE_FIN_PATHO	Date de début	Date inférieure à : INCLUSION.DTE_NAISS, Confirmez-vous?	ANTEC.DTE_FIN_PATHO < INCLUSION.DTE_NAISS

G. Conventions de saisie / Mode d'emploi

- Conventions générales
  - Comment accéder au logiciel de saisie
  - Comment créer, modifier et valider un dossier
  - Codage des données manquantes ou illisibles
    - ND/08/2012 -> 01/01/2012
    - Saisir « ILLISIBLE » dans la zone de texte
    - Cocher « ANOMALIE » si plus d'une case est cochée

- Conventions particulières
  - Conventions spécifiques pour un formulaire

H. Test de la base

- Saisie de patients fictifs
  - Masque de saisie vs CRF papier
  - Convivialité, ergonomie
  - Champs de saisie (taille, formats…)

- Extraction des données
  - Vérification des enregistrements dans les tables
  - Vérification des requêtes d'export

- => **Mise en production**

II. Saisie des données

A. Type de support
B. Réception des CRF papiers
C. Type de saisie
D. Audit-trail
E. Sauvegardes
F. Exemple

A. Type de support

- Cahier d'observation papier (CRF)
  - Saisie dans la base par un opérateur de saisie
  - Saisie faite à partir des CRF papiers

- Cahier d'observation électronique (eCRF)
  - Saisie directe par les investigateurs / TECs
  - Pas de support papier

B. Réception des CRF papiers

- Le Data Manager vérifie :
  - Remplissage du bandeau d'identification du CRF
  - Pages manquantes ou illisibles

- Suivi de la réception des CRF :
  - Nombre de CRF reçus le mois en cours
  - Nombre de CRF reçus au total

C. Type de saisie

- Saisie simple

- Double saisie
  - Indépendante
    - Confrontation des bases 1 et 2 par le Data Manager
    - Correction des erreurs en se basant sur le CRF
  - Interactive
    - En cas de discordance, une alerte est présentée au deuxième opérateur de saisie

C. Type de saisie (2)

- Import des données externes
  - Résultats de laboratoires, imagerie…
- Suivi de la saisie (global et mensuel)
  - Nombre de CRF saisis en base 1
  - Nombre de CRF saisis en base 2

D. Audit-trail

- Suivi des saisies et des modifications des données :
  - **Qui ?** (Nom de l'utilisateur)
  - **Quand ?** (Date de la saisie)
  - **Quoi ?** (variable concernée)
  - **Ancienne valeur ?** (vide si saisie initiale)
  - **Nouvelle valeur ?**

F. Sauvegardes

- Copie de la base sur différents supports (serveurs, disque local…) par le Data Manager ou l'administrateur
- Sauvegardes fréquentes
- Archivage des anciennes versions

Garantie de ne pas perdre le travail

Fabrice Gourmelon - URC/CIC Necker - Cochin

F. Exemple

DEMONSTRATION

Base de données de démonstration

Version---

Créer un patient

Recherche de patients : cliquer sur un ou plusieurs des critères ci-dessous pour activer les filtres

☑ Référence du patient ☑ Centre ☑ Investigateur

Liste des patients : 4 patient(s) sur un total de 4

Référence	Centre	Investigateur
01-001-A-B	01	Dr TESTEUR
01-002-C-D	01	Dr TESTEUR
02-001-E-F	02	Dr DEMONSTRATION
02-002-G-H	02	Dr DEMONSTRATION

Sélection du patient

Double-cliquer sur la ligne correspondant au patient pour accéder aux données

F. Exemple (2)

DEMONSTRATION
Référence : 01-002-C-D
Visite : INCLUSION
PATIENT
Suivi

Suivi

- Données démographique - Consentement - Critères d'inclusion - Critères de non inclusion
- Antécédents médicaux
- Histoire de la maladie
- Visite 1
- Visite 2
- Fin d'étude
- Evénements indésirables (EI)
- Traitements concomitants

Sélection de la visite

QUITTER

F. Exemple (3)

DEMONSTRATION

Référence : 01-002-C-D

Visite : VISITE1

PATIENT

Visite de suivi

Interrogatoire du patient | Examen clinique | Bilan biologique | Bilan biologique (suite)

=> Date du prélèvement :	15/08/2012
=> Hémoglobine (g/dL) :	15
=> Hématocrite (%) :	60
=> VGM (µm³) :	100
=> CCMH (g/dL) :	20
=> TCMH (pg) :	30
=> Réticulocytes (/mm³) :	150000
=> Neutrophiles (/mm³) :	10000
=> Eosinophiles (/mm³) :	800
=> Basophiles (/mm³) :	200

Saisie des données

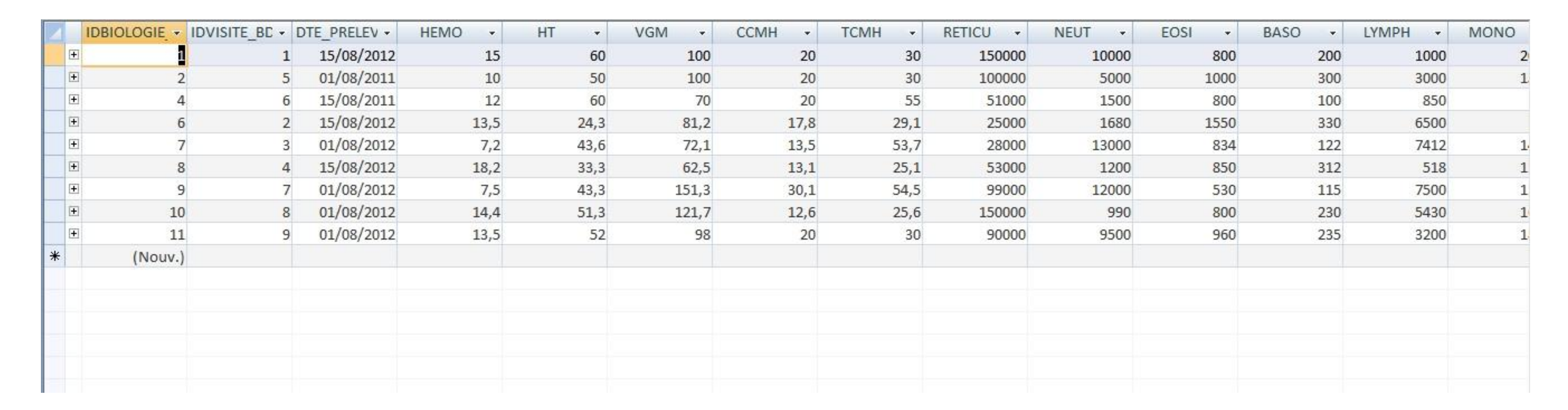
VALIDER

Annuler

II. Saisie des données

Fabrice Gourmelon - URC/CIC Necker - Cochin

F. Exemple (4)



IDBIOLOGIE	IDVISITE_BD	DTE_PRELEV	HEMO	HT	VGM	CCMH	TCMH	RETICU	NEUT	EOSI	BASO	LYMPH	MONO
1	1	15/08/2012	15	60	100	20	30	150000	10000	800	200	1000	2
2	5	01/08/2011	10	50	100	20	30	100000	5000	1000	300	3000	1
4	6	15/08/2011	12	60	70	20	55	51000	1500	800	100	850	
6	2	15/08/2012	13,5	24,3	81,2	17,8	29,1	25000	1680	1550	330	6500	
7	3	01/08/2012	7,2	43,6	72,1	13,5	53,7	28000	13000	834	122	7412	1
8	4	15/08/2012	18,2	33,3	62,5	13,1	25,1	53000	1200	850	312	518	1
9	7	01/08/2012	7,5	43,3	151,3	30,1	54,5	99000	12000	530	115	7500	1
10	8	01/08/2012	14,4	51,3	121,7	12,6	25,6	150000	990	800	230	5430	1
11	9	01/08/2012	13,5	52	98	20	30	90000	9500	960	235	3200	1
(Nouv.)													

Résultat en base de données

III. Contrôle des données

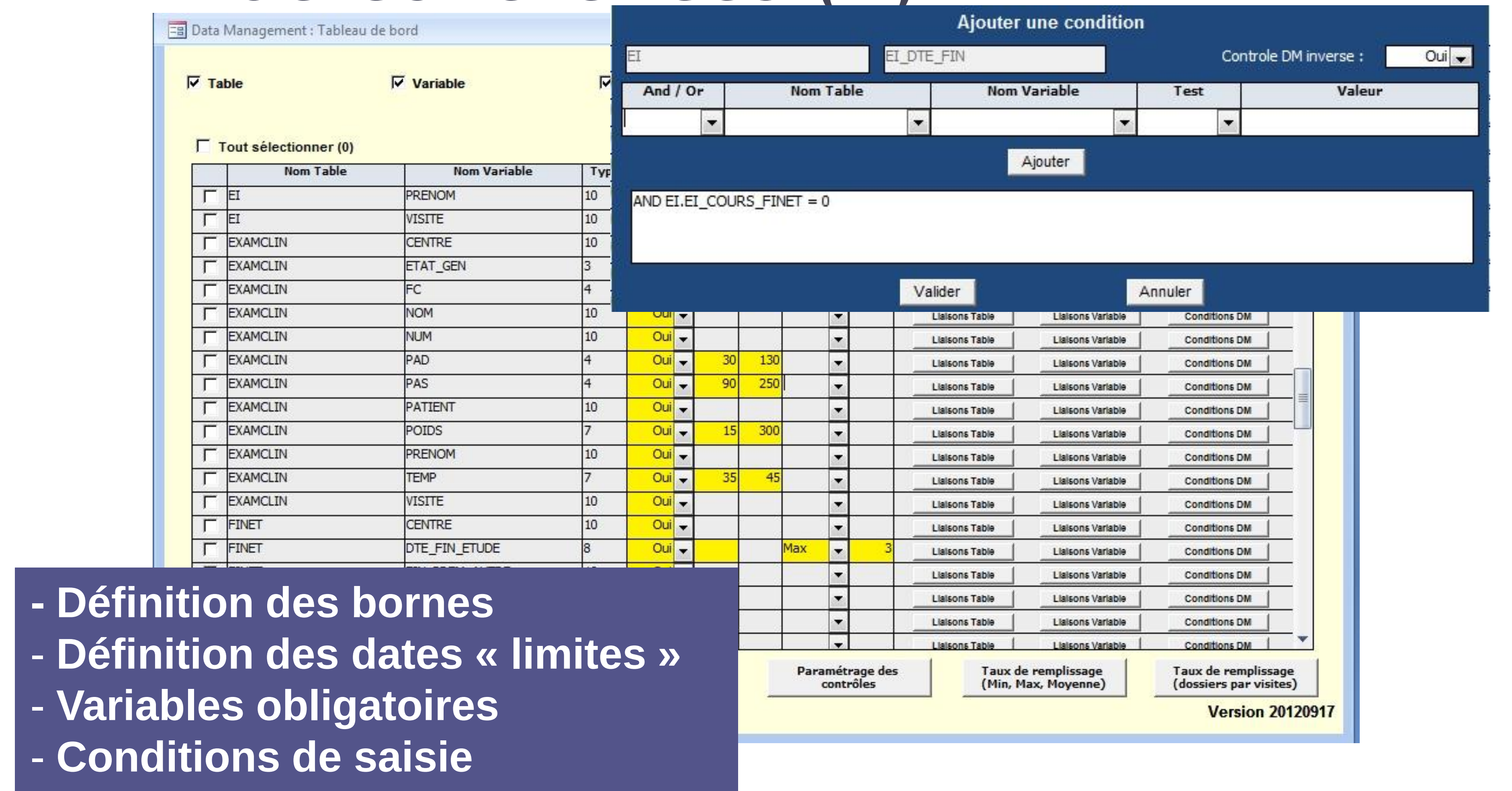
A. Programmation des contrôles de cohérences
B. Edition de listing des queries
C. Circuit des demandes de correction (DCF)
D. Correction des données
E. Statut des queries
F. Description des données
G. Contrôle qualité de la base de données

A. Programmation des contrôles de cohérences

Requêtes SQL détectant les incohérences :

- Déviations au protocole
  - Critères d'inclusion cochés sur « Non »
- Incohérences sur une donnée ou entre plusieurs données
  - Test de grossesse positif pour un homme
  - Date de la visite 2 antérieure à la date de visite 1
- Données manquantes

A. Programmation des contrôles de cohérences (2)



- Définition des bornes
- Définition des dates « limites »
- Variables obligatoires
- Conditions de saisie

A. Programmation des contrôles de cohérences (3)

Data Management : Edit-Check

Paramètres à afficher dans le listing des demandes de c

Référence : EC1 *(ex : EC1, EC2 ...)*

Table : INCLUSION *Table concernée par le contrôle*

Variable : INCL_1 *Variable concernée par le contrôle*

Message : Le critère d'inclusion est coché sur "non"

Intitulé de la demande de correction

Conditions de l'affichage

Source : INCLUSION *Clause "FROM" de la requête (Peut-être une table ou une requête)*

Variable : INCL_1 *Nom de la variable concernée dans la source ci-dessus*

Contrôle : INCLUSION.INCL_1 = 0

Clause "WHERE" de la requête

Etat : Activé

Valider | Annuler | Supprimer

- Incohérences entre les données
- Déviation au protocole

B. Edition des listings des queries

- Exécution des programmes de contrôles
  - Listing des queries ouvertes
    - Référence / Type
    - Date de l'émission
    - Référence du patient
    - Nom de la visite
    - Table et variable de la base de données
    - Valeur de la données
    - Message de la query

B. Edition des listings des queries (2)

	ID	TYPE	DATE	PATIENT	VISITE	NOM_TABLE	LABEL_VARIABLE	VALEUR	QUERY	REPONSE	SIGNATURE
2	8	EC10	12/11/2012	01-001-A-B		VISITE	Date de la visite	01/08/2012	La date de la visite 1 est supérieure à la date de la visite 2, Confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	
3	33	DA1	12/11/2012	01-001-A-B	VISITE1	BIOLOGIE	Date du prélèvement	01/08/2011	Date inférieure à : INCLUSION.DTE_CONSENT, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	
4	13	EC13	12/11/2012	01-001-A-B	TRT CONCOMITANT	TRT_CONC	Date de début	10/08/2012	La date de début du traitement est supérieure à la date de fin, Confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	
5	35	DA1	12/11/2012	01-001-A-B	HISTOIRE MALADIE	HISTMED	Date de début des symptômes	01/01/1949	Date inférieure à : INCLUSION.DTE_NAISS, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	
6	36	DA1	12/11/2012	01-001-A-B	TRT CONCOMITANT	TRT_CONC	Date de début	15/08/2011	Date inférieure à : INCLUSION.DTE_CONSENT, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	
7	1	EC1	12/11/2012	01-001-A-B	INCLUSION	INCLUSION	Patient âgé de plus de 18 ans	0	Le critère d'inclusion est coché sur "non", Confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :	

C. Circuit des demandes de corrections (DCF)

Incohérences => Demande de correction (DCF)

Sauf pour les corrections autorisées

- Transmission des DCF
  - Data Manager => ARC => Investigateur
- Retour des DCF
  - Investigateur => ARC => Data Manager
- Suivi des envois et retours
  - Nombre d’envois et de retours le mois en cours
  - Nombre d’envois et de retours au total

C. Circuit des demandes de corrections (DCF) (2)

FICHE DE CORRECTIONS DE DONNEES

Page 14 sur 14

Protocole : ETUDE   Patient : 03-001-I-J   Date d'émission : 12/11/2012 00:48:38

ID	TYPE	DATE	VISITE	TABLE	VARIABLE	VALEUR	QUERY	REPONSE
24	BO1	12/11/2012	VISITE1	BIOLOGIE	Bicarbonates (mmol/L)	2	Donnée hors bornes [10 - 80], confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :
40	DM1	12/11/2012	ANTECEDENTS	ANTEC	Date de début	null	Donnée manquante, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :
50	DM1	12/11/2012	VISITE1	EXAMCLIN	Poids (Kg)	null	Donnée manquante, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :
57	DM2	12/11/2012	ANTECEDENTS	ANTEC	Date de début	01/01/2009	Donnée valorisée alors qu'aucune réponse n'est attendue, confirmez-vous ?	1 - Confirmer la donnée 2 - Donnée non vérifiable 3 - Corriger la donnée :

Date, Nom et signature de l'investigateur :   Date, Nom et signature du Data Manager :   Date de réception URC :

D. Correction des données

- Retour des DCF
  - Saisie des nouvelles valeurs dans la base à partir des DCF datées et signées par l'investigateur

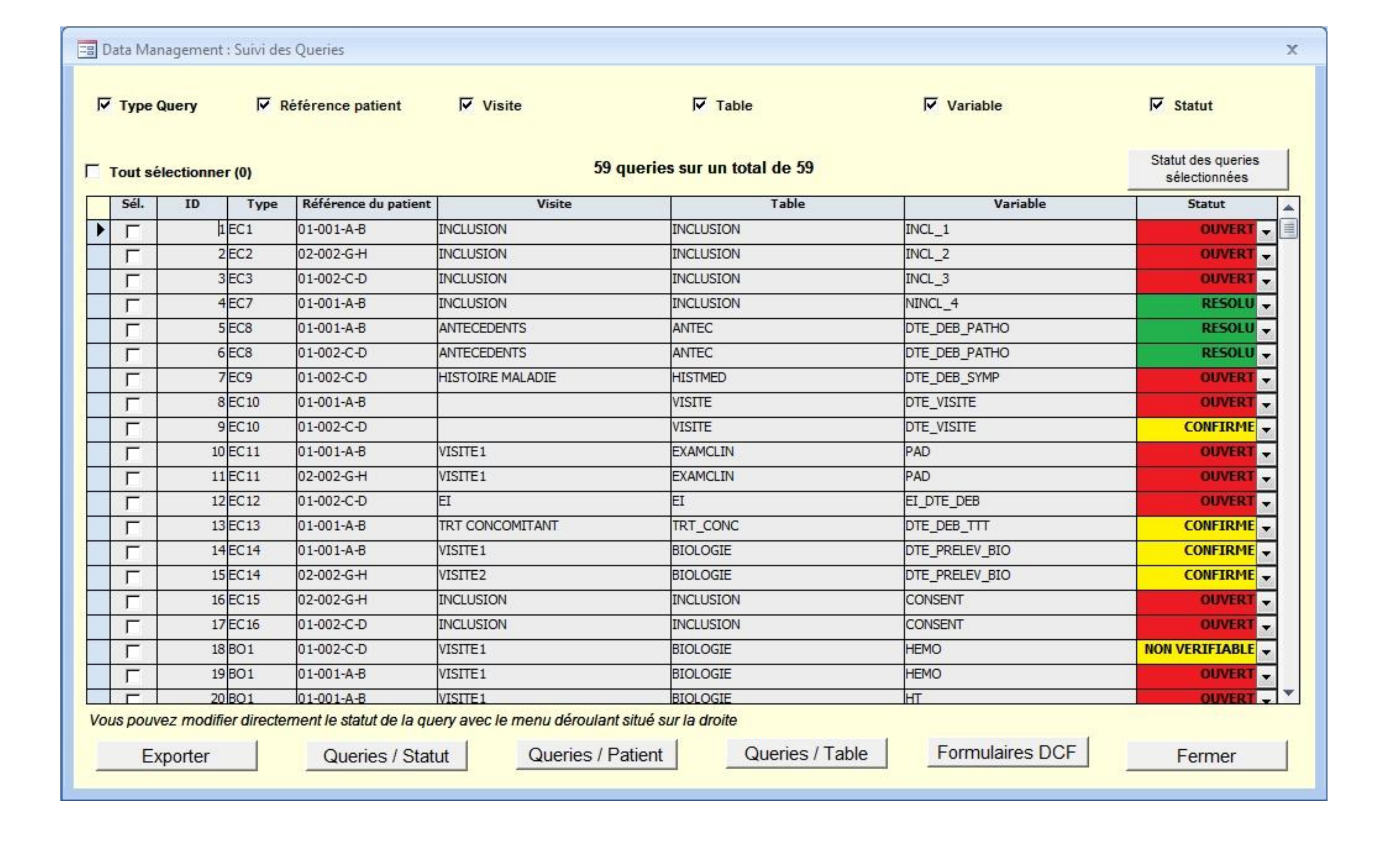
- Corrections autorisées
  - Correction des données par le Data Manager sans passer par une DCF
  - Les corrections autorisées sont définies
    - Traitements pris : ☐Oui ☒Non
      - Si oui, préciser : Aspirine

E. Statut des queries

- Statut des queries :
  - Ouvert
  - Résolu
  - Confirmé
  - Non vérifiable

- Exécution régulières des contrôles et émission des DCF
  - Tant que la saisie n'est pas terminée
  - Tant qu'il reste des queries ouvertes

E. Statut des queries (2)



Data Management : Suivi des Queries

☑ Type Query ☑ Référence patient ☑ Visite ☑ Table ☑ Variable ☑ Statut

☐ Tout sélectionner (0)

59 queries sur un total de 59

Statut des queries sélectionnées

Sél.	ID	Type	Référence du patient	Visite	Table	Variable	Statut
☐	1	EC1	01-001-A-B	INCLUSION	INCLUSION	INCL_1	OUVERT
☐	2	EC2	02-002-G-H	INCLUSION	INCLUSION	INCL_2	OUVERT
☐	3	EC3	01-002-C-D	INCLUSION	INCLUSION	INCL_3	OUVERT
☐	4	EC7	01-001-A-B	INCLUSION	INCLUSION	NINCL_4	RESOLU
☐	5	EC8	01-001-A-B	ANTECEDENTS	ANTEC	DTE_DEB_PATHO	RESOLU
☐	6	EC8	01-002-C-D	ANTECEDENTS	ANTEC	DTE_DEB_PATHO	RESOLU
☐	7	EC9	01-002-C-D	HISTOIRE MALADIE	HISTMED	DTE_DEB_SYMP	OUVERT
☐	8	EC10	01-001-A-B		VISITE	DTE_VISITE	OUVERT
☐	9	EC10	01-002-C-D		VISITE	DTE_VISITE	CONFIRME
☐	10	EC11	01-001-A-B	VISITE1	EXAMCLIN	PAD	OUVERT
☐	11	EC11	02-002-G-H	VISITE1	EXAMCLIN	PAD	OUVERT
☐	12	EC12	01-002-C-D	EI	EI	EI_DTE_DEB	OUVERT
☐	13	EC13	01-001-A-B	TRT CONCOMITANT	TRT_CONC	DTE_DEB_TTT	CONFIRME
☐	14	EC14	01-001-A-B	VISITE1	BIOLOGIE	DTE_PRELEV_BIO	CONFIRME
☐	15	EC14	02-002-G-H	VISITE2	BIOLOGIE	DTE_PRELEV_BIO	CONFIRME
☐	16	EC15	02-002-G-H	INCLUSION	INCLUSION	CONSENT	OUVERT
☐	17	EC16	01-002-C-D	INCLUSION	INCLUSION	CONSENT	OUVERT
☐	18	BO1	01-002-C-D	VISITE1	BIOLOGIE	HEMO	NON VERIFIABLE
☐	19	BO1	01-001-A-B	VISITE1	BIOLOGIE	HEMO	OUVERT
☐	20	BO1	01-001-A-B	VISITE1	BIOLOGIE	HT	OUVERT

Vous pouvez modifier directement le statut de la query avec le menu déroulant situé sur la droite

Exporter | Queries / Statut | Queries / Patient | Queries / Table | Formulaires DCF | Fermer

F. Description des données

- Taux de remplissage pour chaque données
  - Nombre de données saisies
  - Nombre de données attendues

- Statistique descriptives pour les données quantitatives
  - Minimum
  - Maximum
  - Moyenne

F. Description des données (2)

	A	B	C	D	E	F	G	H	I
1	TABLE	VARIABLE	TYPE	NB_ATTENDU	NB_SAISIE	TAUX_SAISIE	V_MIN	V_MAX	V_MOY
2	ANTEC	CENTRE	10	8	8	100			
3	ANTEC	COURS_PATHO	3	8	8	100			
4	ANTEC	DTE_DEB_PATHO	8	7	7	100			
5	ANTEC	DTE_FIN_PATHO	8	8	6	75			
6	ANTEC	NOM	10	8	8	100			
7	ANTEC	NOM_PATHO	10	8	8	100			
8	ANTEC	NUM	10	8	8	100			
9	ANTEC	NUM_PATHO	4	8	8	100	1	2	1,38
10	ANTEC	PATIENT	10	8	8	100			
11	ANTEC	PRENOM	10	8	8	100			
12	ANTEC	VISITE	10	8	8	100			
13	BIOLOGIE	ALAT	4	9	9	100	28	80	48,22
14	BIOLOGIE	ASAT	4	9	8	89	12	75	42
15	BIOLOGIE	BASO	4	9	9	100	100	330	216
16	BIOLOGIE	CA	7	9	9	100	1,9	33	7,74
17	BIOLOGIE	CCMH	7	9	8	89	12,6	30,1	18,39
18	BIOLOGIE	CENTRE	10	9	9	100			
19	BIOLOGIE	CL	7	9	9	100	10	130,2	88,5
20	BIOLOGIE	CREAT	7	9	9	100	38	500	145,77
21	BIOLOGIE	DTE_PRELEV_BIO	8	9	9	100			
22	BIOLOGIE	EOSI	4	9	7	78	800	1550	942
23	BIOLOGIE	GGT	4	9	9	100	16	80	51,56
24	BIOLOGIE	HCO3	7	9	8	89	2	50	28,75
25	BIOLOGIE	HEMO	7	9	9	100	7,2	250	40,7
26	BIOLOGIE	HT	7	9	9	100	24,3	158	58,42
27	BIOLOGIE	K	7	9	9	100	2,1	4,3	3,37
28	BIOLOGIE	LYMPH	4	9	9	100	518	7500	3934,44
29	BIOLOGIE	MONO	4	9	9	100	75	2000	1222,33
30	BIOLOGIE	NA	7	9	9	100	10	218	113,62
31	BIOLOGIE	NEUT	4	9	8	89	990	13000	5358,75
32	BIOLOGIE	NOM	10	9	9	100			
33	BIOLOGIE	NUM	10	9	9	100			
34	BIOLOGIE	PATIENT	10	9	9	100			
35	BIOLOGIE	PLAQ	4	9	9	100	90000	426000	262333,33

G. Contrôle qualité de la base de données

- Comparaison : CRF papier vs base de données
- Tirage au sort des CRF papier
  - RACINE(N) + 1 (N = Nombre de CRF)
- Vérification de 100% des items du critère de jugement principal
  - Taux d'erreur accepté : 1 pour 1000

G. Contrôle qualité de la base de données (2)

- Test concluant (taux d’erreur < 1/1000)
  - Données de bonne qualité

- Test non concluant (taux d’erreur > 1/1000)
  - On refait un tirage au sort
    - Test concluant (taux d’erreur < 1/1000)
      - Données de bonne qualité
    - Test non concluant (taux d’erreur > 1/1000)
      - La saisie doit être refaite

IV. Codification

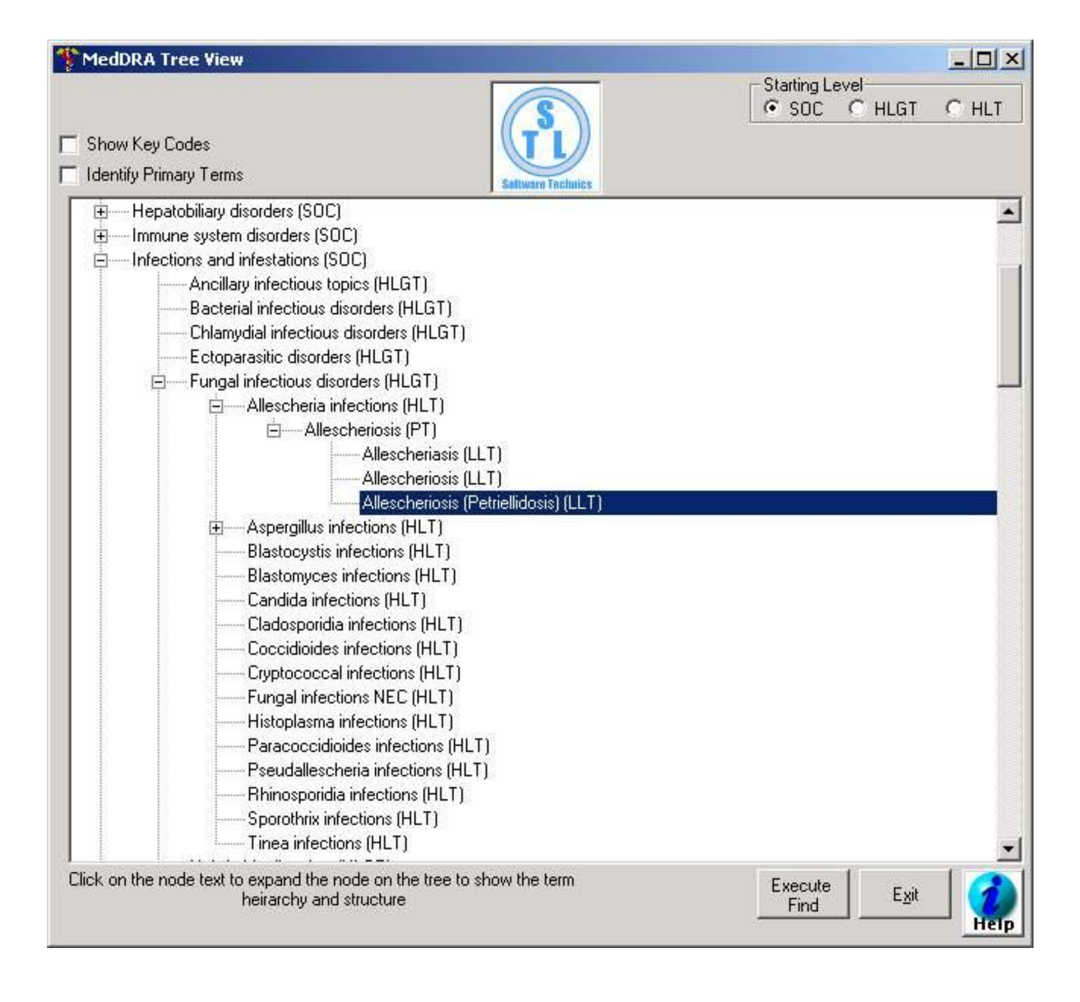
A. Définition
B. MedDRA

A. Définition

- Choix d'un terme de référence en vue d'une exploitation informatique

- Que doit-on coder ?
  - Les événements indésirables
  - Les traitements médicamenteux
  - Les actes diagnostiques
  - Les actes thérapeutiques

B. MedDRA


MedDRA Tree View
Starting Level
SOC
HLGT
HLT
Show Key Codes
Identify Primary Terms
Software Technics
Hepatobiliary disorders (SOC)
Immune system disorders (SOC)
Infections and infestations (SOC)
Ancillary infectious topics (HLGT)
Bacterial infectious disorders (HLGT)
Chlamydial infectious disorders (HLGT)
Ectoparasitic disorders (HLGT)
Fungal infectious disorders (HLGT)
Allescheria infections (HLT)
Allescheriosis (PT)
Allescheriasis (LLT)
Allescheriosis (LLT)
Allescheriosis (Petriellidosis) (LLT)
Aspergillus infections (HLT)
Blastocystis infections (HLT)
Blastomyces infections (HLT)
Candida infections (HLT)
Cladosporidia infections (HLT)
Coccidioides infections (HLT)
Cryptococcal infections (HLT)
Fungal infections NEC (HLT)
Histoplasma infections (HLT)
Paracoccidioides infections (HLT)
Pseudallescheria infections (HLT)
Rhinosporidia infections (HLT)
Sporothrix infections (HLT)
Tinea infections (HLT)
Click on the node text to expand the node on the tree to show the term heirarchy and structure
Execute Find
Exit
Help

V. Réconciliation avec la base de PV

Réconciliation avec la base de PV

- Echange d'information sur les EIG avec la PV du promoteur
- Vérification des EIG en commun
  - Patient profile : caractéristiques + données EIG
  - Nombre d'EIG dans chaque base
- En cas de différence :
  - Requête de PV

VI. Blind-Review

A. Définition
B. Objectifs

A. Définition

- Réunion de revue des données : c'est la mise en commun de toutes les information de l'essai
- Elle a lieu quand toutes les données de l'étude sont recueillies, saisies et validées
- Protagonistes : Data Manager, Chef de Projets (+/- ARC), Statisticien, Investigateur coordonnateur

B. Objectifs

- Revoir les conditions de réalisation de l’étude
- Spécifier les déviations définies dans le protocole
- Vérifier les paramètres d’efficacité
- Déterminer les population d’analyse
- Revoir le plan d’analyse statistique

Des requêtes peuvent encore être émises suite à cette réunion

VII. Gel de la base / Transfert

A. Gel de la base
B. Transfert des données au statisticien

A. Gel de la base

- Tous les CRF sont saisis
- Toutes les corrections sont intégrées
- Monitoring terminé
- Contrôle de cohérences : pas d'anomalies
- Codification validée
- Contrôle qualité effectué
- Consolidation avec la PV
- Réunion de Blind-Review réalisée
- Plan d'analyse statistique validé

A. Gel de la base (2)

- Verrouillage informatique par le Data Manager:
  - Les données ne sont pas modifiables (accessibles en lecture seule)

DEMONSTRATION

Référence : 01-001-A-B

Visite : VISITE1

PATIENT

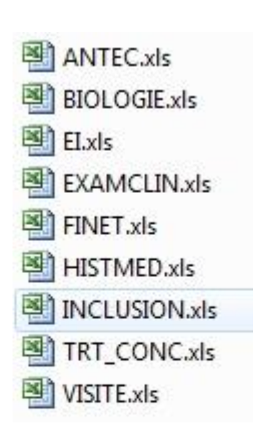
Visite de suivi

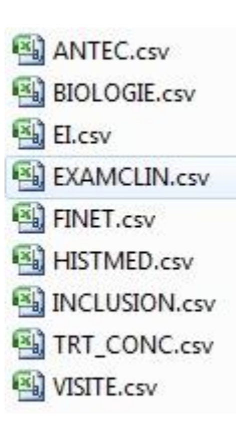
Interrogatoire du patient | Examen clinique | Bilan biologique | Bilan biologique (suite)

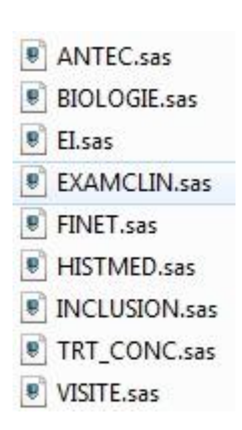
=> Date du prélèvement :	01/08/2011
=> Hémoglobine (g/dL) :	30
=> Hématocrite (%) :	158
=> VGM (µm³) :	100

B. Transfert

- Les fichiers de la base de données et la documentation sont transférés au statisticien pour analyse :
  - Protection par un mot de passe
  - Format des fichiers (SAS, Excel, CSV…)



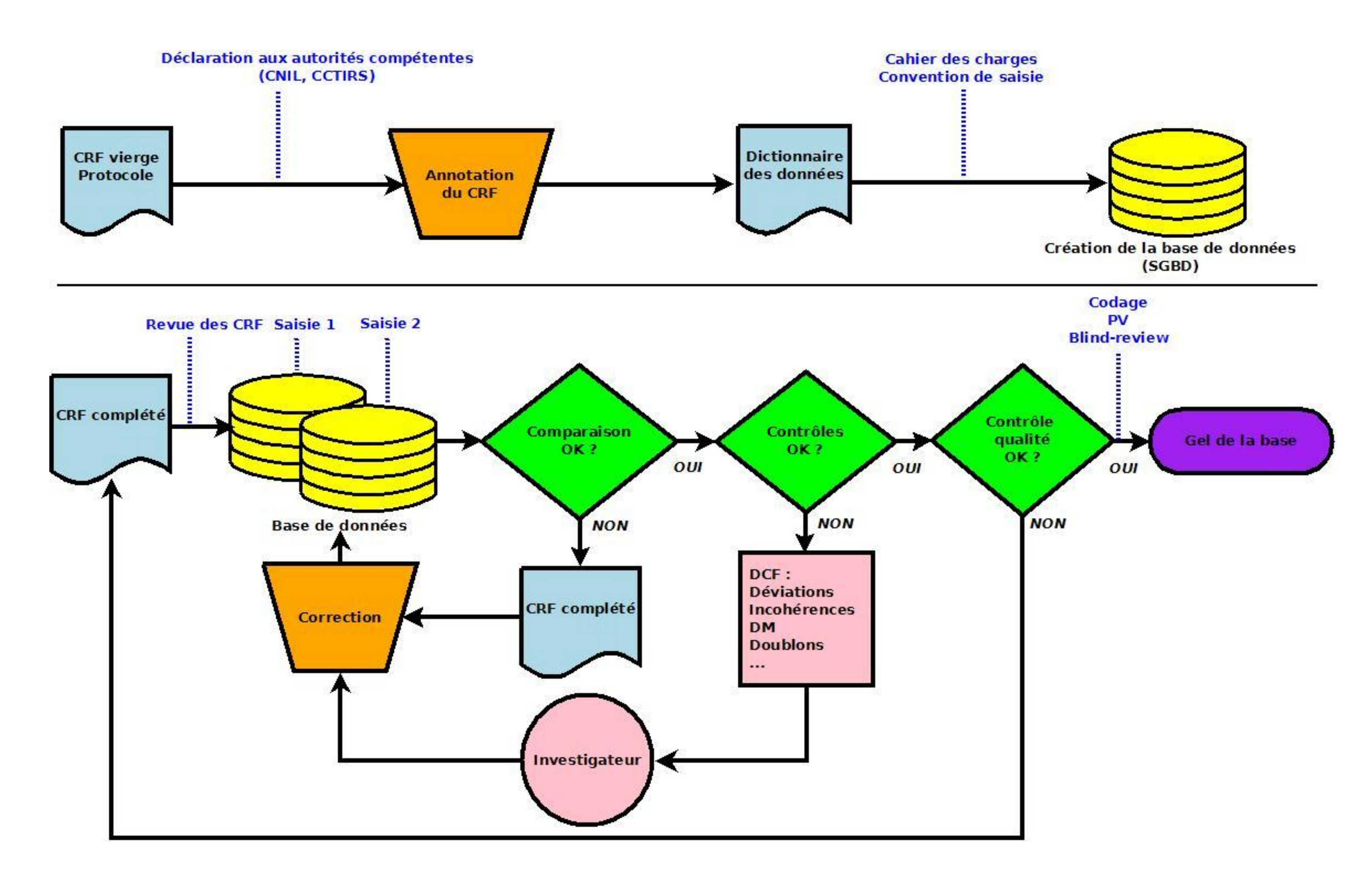




Conclusion

A. Etapes de data management
B. Conseils

A. Etapes de data management



B. Conseils

- Impliquer tous les intervenants dans la préparation
  - CP, Investigateur, ARC, TEC, Data Manager, Biostatisticien
- Ne pas « brûler » les étapes
  - Le temps « gagné » au début sera perdu à la fin
- Modifications du CRF en cours d'étude
  - Impact sur toutes les étapes du data management
  - Perte de temps
  - Risque (modifications sur une base en production)